

Package ‘metaITH’

Elise Merritt

2019-06-26

Package: metaITH

Title: A Tool for Analyzing Intratumor Heterogeneity at Multiple Levels

Version: 0.0.0.9000

Author: Elise Merritt, em575@scarletmail.rutgers.edu

Description: This package allows for analysis of intratumor heterogeneity at the DNA, RNA, and immunogenic levels.

Depends: R ($\geq 3.5.0$), ggplot2, reshape2, ape, phylobase, dplyr, matrixStats

License: MIT

Encoding: UTF-8

LazyData: true

RoxygenNote: 6.1.1

Suggests: knitr, rmarkdown

VignetteBuilder: knitr

R topics documented:

| | |
|---|-----------|
| Signatures | 2 |
| z_score_calculations | 2 |
| anti_pd1_favor_signature | 2 |
| apoptosis_signature | 3 |
| drug_resistance_signature | 4 |
| emt_signature | 5 |
| hypoxia_signature | 6 |
| proliferation_signature | 6 |
| specified_geneset_signature | 7 |
| signature_analysis | 8 |
| Phylograms | 10 |
| dna_phylograms | 10 |
| immune_phylograms | 10 |
| rna_phylograms | 11 |
| snv_heatmaps | 12 |
| Multi-region Divergence and Diversity | 12 |
| multi_level_divergence_diversity | 13 |
| Utilities | 14 |
| metaITH_analysis | 14 |
| Index | 17 |

Signatures

z_score_calculations

Description

This function must be run before doing any signature analysis function. The input is an RNA expression matrix containing the expression level of all genes in all tumor samples and normals. The input file must be placed manually in metaITH/R directory. The output is a matrix (tab-delimited text file) containing z-scores for all genes in all of the tumor samples or regions and normals. NOTE: All normal samples must end in “.nr” and all tumor samples must not end in “.nr”

Usage

```
z_score_calculations(rna_expression_matrix_file)
```

Arguments

- `rna_expression_matrix_file`: File containing expression level of all genes in all tumor samples. Header of first column should be “Gene” and the header of any column with data from normal samples should end in “.nr”.

Details

This function takes in RNA expression matrix containing the expression level of all genes in all tumor samples and normals and creates a matrix (tab-delimited text file) containing z-scores for all genes in all of the tumor samples or regions and normals.

Value

1. A table of z-score values. First column contains all genes found in RNA expression matrix. Subsequent columns contain z-score of each gene for each sample.

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
z_score_calculations("RNA_matrix.txt")
read.table("z-scores_matrix.txt", sep="\t", header=T)[1:4,1:4]
```

```
##      Gene  brs1.t1r.  brs1.t2r.  brs1.t3r.
## 1 RN7SL1  0.5900522  0.58684943 -0.08013765
## 2 SFTPC   -0.3472531 -1.98528401  0.09348880
## 3 AKR1C1  -0.8676535 -0.28814552 -0.74276365
## 4 AKR1C2  -0.7285173 -0.06270846 -0.69599408
```

anti_pd1_favor_signature

Description

The input file is a z-score matrix created by function `z_score_calculations`. The input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. The function outputs a table of anti-PD1 favor scores for each tumor sample or region and a heatmap of the anti-PD1 favor scores.

Usage

```
anti_pd1_favor_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created by running the function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of anti-PD1 favor scores for each tumor sample or region and a heatmap of the anti-PD1 favor scores.

Value

1. A table of anti-PD1 favor scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.
2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
anti_pd1_favor_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("anti-PD1_favor_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 anti-PD1 favor score 0.473427 0.1254103 0.7149302
```

apoptosis_signature

Description

Input is a z-score matrix created by function `z_score_calculations`. Input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. Outputs a table of apoptosis scores for each tumor sample or region and a heatmap of the apoptosis scores.

Usage

```
apoptosis_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created by running the function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of apoptosis scores for each tumor sample or region and a heatmap of the apoptosis scores.

Value

1. A table of apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.

2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
apoptosis_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Apoptosis_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Apoptosis score 0.5290366 1.116061 0.8882319
```

drug_resistance_signature

Description

Input is a z-score matrix created by function `z_score_calculations`. Input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. Outputs a table of drug-resistance scores for each tumor sample or region and a heatmap of the drug-resistance scores.

Usage

```
drug_resistance_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created by running function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of drug-resistance scores for each tumor sample or region and a heatmap of the drug-resistance scores.

Value

1. A table of drug-resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Pemetrexed resistance score” shows the drug-resistance score for each tumor region or sample.
2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
drug_resistance_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Pemetrexed_resistance_score.txt", sep="\t", header=T)[,1:4]
```

```
##                               Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Pemetrexed resistance score 0.4137076  1.293488  0.814833
```

emt_signature

Description

Input is a z-score matrix created by function `z_score_calculations`. Input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. Outputs a table of emt(epithelial-mesenchymal transition) scores for each tumor sample or region and a heatmap of the emt scores.

Usage

```
emt_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created using function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of emt(epithelial-mesenchymal transition) scores for each tumor sample or region and a heatmap of the emt scores.

Value

1. A table of emt scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
2. A table of mesenchymal-epithelial scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
3. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
emt_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Mesenchymal-Epithelial_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 M-E score 0.5018744 -0.3082992 0.05229863
```

hypoxia_signature

Description

Input is a z-score matrix created by function `z_score_calculations`. Input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. Outputs a table of hypoxia scores for each tumor sample or region and a heatmap of the hypoxia scores.

Usage

```
hypoxia_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created using function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of hypoxia scores for each tumor sample or region and a heatmap of the hypoxia scores.

Value

1. A table of hypoxia scores. First column is "Gene", followed by one column for each tumor region or sample. Two rows, one for sample names, one called "Hypoxia score" shows the hypoxia score for each tumor region or sample.
2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")  
hypoxia_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to  
## character vector
```

```
## pdf
```

```
## 2
```

```
read.table("Hypoxia_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.  
## 1 Hypoxia score 0.4360642  2.036502  0.570733
```

proliferation_signature

Description

Input is a z-score matrix created by function `z_score_calculations`. Input file must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. Outputs a table of proliferation scores for each tumor sample or region and a heatmap of the proliferation scores.

Usage

```
proliferation_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. This file should be created using function `z_score_calculations`.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of proliferation scores for each tumor sample or region and a heatmap of the proliferation scores.

Value

1. A table of proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
proliferation_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478
```

specified_geneset_signature

Description

Takes in a z-score matrix created by function `z_score_calculations` and a geneset. Input file z-core matrix must be in metaITH/R directory, which is automatically done by function `z_score_calculations`. The geneset must be placed in the metaITH/R directory manually. Outputs a table of geneset scores and a geneset scores heatmap.

Usage

```
specified_geneset_signature(z_score_matrix_file, geneset)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Can be created using function `z_score_calculations`.
- `geneset`: A text file containing a list of genes supplied by the user, one gene per line.

Details

Takes in a z-score matrix created by function `z_score_calculations` and a geneset. Outputs a table of geneset scores and a geneset scores heatmap.

Value

1. A table of geneset scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Geneset” score” shows the geneset score for each

tumor region or sample.

2. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
specified_geneset_signature("z-scores_matrix.txt", "proliferation_geneset.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Geneset_score.txt", sep="\t", header=T)[,1:4]
```

```
##          Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Geneset score 0.09291747 0.8304992 0.4802478
```

signature_analysis

Description

Input is a RNA expression matrix containing the expression level of all genes in all tumor samples and normals. NOTE: All normal samples should end in “nr” and all tumor samples should not end in “nr”. Input file must be placed manually in metaITH/R directory. Creates a z-score matrix and uses this matrix to find the hypoxia, proliferation, apoptosis, drug-resistance, emt, and anti-PD1 favor scores for each tumor sample or region. Outputs a table and heatmap of these scores.

Usage

```
signature_analysis(rna_expression_matrix_file)
```

Arguments

- `rna_expression_matrix_file`: File containing expression level of all genes in all tumor samples. Header of first column should be “Gene” and each normal file should end in “nr”

Details

Takes in RNA expression matrix containing the expression level of all genes in all tumor samples and normals. Creates a z-score matrix and uses this matrix to find the hypoxia, proliferation, apoptosis, drug-resistance, emt, and anti-PD1 favor scores for each tumor sample or region. Outputs a table and heatmap for each of these scores.

Value

1. A table of hypoxia scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Hypoxia score” shows the hypoxia score for each tumor region or sample.
2. A tiff formatted image file visualizing the hypoxia scores
3. A table of proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
4. A tiff formatted image file visualizing the proliferation scores

5. A table of apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.
6. A tiff formatted image file visualizing the apoptosis scores
7. A table of drug-resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Pemetrexed resistance score” shows the drug-resistance score for each tumor region or sample.
8. A tiff formatted image file visualizing the drug-resistance scores
9. A table of emt scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
10. A table of mesenchymal-epithelial scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
11. A tiff formatted image file visualizing the mesenchymal-epithelial scores
12. A table of anti-PD1 favor scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.
13. A tiff formatted image file visualizing the anti-PD1 scores

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
signature_analysis("RNA_matrix.txt")

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## pdf
## 2

read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene   brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478
```

Phylograms

dna_phylograms

Description

Takes in a text file containing list of sample files. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal. The input text file and all the sample files must be placed in the metaITH/R directory. Creates a DNA distance matrix, containing the distance between all the tumor regions and normal, a DNA tree topology file, and an unrooted DNA phylogram.

Usage

```
dna_phylograms(dna_phylo_list_file)
```

Arguments

- `dna_phylo_list_file`: A text file containing a list of sample files, one file name per line. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal.

Details Takes in a text file containing list of sample files. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal. Creates a DNA distance matrix, containing the distance between all the tumor regions and normal, a DNA tree topology file, and an unrooted DNA phylogram.

Value

1. DNA distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.
2. A text file containing a DNA distance tree
3. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
dna_phylograms("DNA_phylo_list.txt")
read.table("DNA_distance_matrix_DNA_S1_frequency_matrix.txt", sep="\t", header=T)
```

```
##           R1           R2           R3           N
## R1  0.000000  4.814473  4.144641  15.52719
## R2  4.814473  0.000000  6.428683  16.06559
## R3  4.144641  6.428683  0.000000  13.74008
## N   15.527186 16.065588 13.740077  0.00000
```

immune_phylograms

Description

Takes in a text file containing list of sample files. The input text file and all the sample files must be placed in the metaITH/R directory. The function then uses the data in the text file to create an immune distance matrix, containing the distance between all the tumor regions and normal, an immune tree topology file, and an unrooted phylogram.

Usage

```
immune_phylograms(immune_phylo_list_file)
```

Arguments

- immune_phylo_list_file: A text file containing a list of sample files, one file name per line. Each file contains a matrix with proportion of immune cells (inferred by CIBERSORT) of all immune cell types in each tumor region and normal.

Details

Takes in a text file containing list of sample files. The function then uses the data in the text file to create an immune distance matrix, containing the distance between all the tumor regions and normal, an immune tree topology file, and an unrooted phylogram.

Value

1. Immune distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.
2. A text file containing a immune distance tree
3. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
immune_phylograms("Immune_phylo_list.txt")
read.table("Immune_distance_matrix_S1_Immune_CIBERSORT_matrix.txt", sep="\t", header=T)
```

| ## | | R1 | R2 | R3 | N |
|----|----|------------|------------|------------|------------|
| ## | R1 | 0.00000000 | 0.02679035 | 0.02137600 | 0.04413436 |
| ## | R2 | 0.02679035 | 0.00000000 | 0.04783462 | 0.06250866 |
| ## | R3 | 0.02137600 | 0.04783462 | 0.00000000 | 0.02289465 |
| ## | N | 0.04413436 | 0.06250866 | 0.02289465 | 0.00000000 |

rna_phylograms

Description

Takes in a text file containing list of sample files. The input text file and all the sample files must be placed in the metaITH/R directory. Creates a RNA distance matrix, containing the distance between all the tumor regions and normal, a RNA tree topology file, and a RNA unrooted phylogram.

Usage

```
rna_phylograms(rna_phylo_list_file)
```

Arguments

- rna_phylo_list_file: A text file containing a list of sample files, one file name per line. Each file contains a matrix with expression values (preferably $\log_2(\text{TPM}+1)$) of all the genes in each tumor region and normal.

Details

Takes in a text file containing list of sample files. Creates a RNA distance matrix, containing the distance between all the tumor regions and normal, a RNA tree topology file, and a RNA unrooted phylogram.

Value

1. RNA distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.
2. A text file containing a RNA distance tree
3. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
rna_phylograms("RNA_phylo_list.txt")
read.table("RNA_distance_matrix_S1_RNA_expresion_matrix.txt", sep="\t", header=T)
```

```
##           R1           R2           R3           N
## R1 0.0000000 0.6834438 0.409917 0.6300742
## R2 0.6834438 0.0000000 0.756774 1.2733578
## R3 0.4099170 0.7567740 0.000000 0.6876690
## N  0.6300742 1.2733578 0.687669 0.0000000
```

snv_heatmaps

Description

Takes in text file containing list of sample files. The input text file and all the sample files must be placed in the metaITH/R directory. Outputs a heatmap of variant allele frequency (vaf) of all variations for each sample.

Usage

```
snv_heatmaps(dna_phylo_list_file)
```

Arguments

- dna_phylo_list_file: A text file containing a list of sample files, one file name per line. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal.

Details

Takes in text file containing list of sample files.

Outputs a heatmap of variant allele frequency (vaf) of all variations for each sample.

Value

1. A tiff formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
snv_heatmaps("DNA_phylo_list.txt")
```

Multi-region Divergence and Diversity

multi_level_divergence_diversity

Description

Takes in text file containing list of samples, one sample per line, and distance matrix files generated by `dna_phylograms`, `rna_phylograms`, and `immune_phylograms` methods. The input text file and all the sample files must be placed in the metaITH/R directory. the distance matrix file will be automatically placed in the metaITH/R directory by the functions `dna_phylograms`, `rna_phylograms`, and `immune_phylograms`. Outputs a png file containing 6 plots, 3 diversity/divergence plots and 3 diversity plots.

Usage

```
multi_level_divergence_diversity(sample_names_file)
```

Arguments

- `sample_names_file`: A text file containing list of sample names, one file name per line

Details Takes in text file containing list of samples, one sample per line, and distance matrix files generated by `dna_phylograms`, `rna_phylograms`, and `immune_phylograms` methods. Outputs a png file containing 6 plots, 3 diversity/divergence plots and 3 diversity plots. NOTE: Requires output of functions `dna_phylograms`, `rna_phylograms`, and `immune_phylograms` to be in same folder.

Value

1. A png formatted image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
multi_level_divergence_diversity("sample_names.txt")
```

```
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
## pdf
## 2
```

Utilities

metaITH_analysis

Description

Analyzes several sample files by running a phylogram analysis (using DNA, RNA, and immune expression data), signature analysis (using RNA expression data), and multi-region diversity and divergence analysis (using DNA, RNA, and immune expression data). All input files must be placed in metaITH/R directory.

Usage

metaITH_analysis(dna_phylo_list, rna_phylo_list, immune_phylo_list, multi_region_sample_names, rna_expression_matrix)

Arguments

- dna_phylo_list: A text file containing a list of sample files, one file name per line. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal.
- rna_phylo_list: A text file containing a list of sample files, one file name per line. Each file contains a matrix with expression values (preferably $\log_2(\text{TPM}+1)$) of all the genes in each tumor region and normal.
- immune_phylo_list: A text file containing a list of sample files, one file name per line. Each file contains a matrix with proportion of immune cells (inferred by CIBERSORT) of all immune cell types in each tumor region and normal.
- multi_region_sample_names: A text file containing list of sample names for multi-region diversity and divergence analysis, one file name per line.
- rna_expression_matrix: File containing expression level of all genes in all tumor samples. Header of first column should be “Gene” and each normal file should end in “.nr”

Details

Analyzes several sample files by running a phylogram analysis (using DNA, RNA, and immune expression data), signature analysis (using RNA expression data), and multi-region diversity and divergence analysis (using DNA, RNA, and immune expression data).

Value

1. DNA distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.
2. A text file containing a DNA distance tree
3. A tiff formatted image file showing a DNA phylogram
4. RNA distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.
5. A text file containing a RNA distance tree
6. A tiff formatted image file showing a RNA phylogram
7. immune distance matrix is a table with one column and one row for each tumor sample or region. Each entry in the table corresponds to the distance between the row sample and column sample. The distance between a sample and itself is 0.

8. A text file containing an immune distance tree
9. A tiff formatted image file showing an immune phylogram
10. A png formatted image file containing 6 plots, 3 diversity/divergence plots and 3 diversity plots
11. A table of hypoxia scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Hypoxia score” shows the hypoxia score for each tumor region or sample.
12. A tiff formatted image file visualizing the hypoxia scores
13. A table of proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
14. A tiff formatted image file visualizing the proliferation scores
15. A table of apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.
16. A tiff formatted image file visualizing the apoptosis scores
17. A table of drug-resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Pemetrexed resistance score” shows the drug-resistance score for each tumor region or sample.
18. A tiff formatted image file visualizing the drug-resistance scores
19. A table of emt scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
20. A table of mesenchymal-epithelial scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
21. A tiff formatted image file visualizing the mesenchymal-epithelial scores
22. A table of anti-PD1 favor scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.
23. A tiff formatted image file visualizing the anti-PD1 scores

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
metaITH_analysis("DNA_phylo_list.txt", "RNA_phylo_list.txt", "Immune_phylo_list.txt", "sample_names.txt")

##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
```

```

## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## pdf
## 2

read.table("DNA_distance_matrix_DNA_S1_frequency_matrix.txt", sep="\t", header=T)

##           R1           R2           R3           N
## R1  0.000000  4.814473  4.144641 15.52719
## R2  4.814473  0.000000  6.428683 16.06559
## R3  4.144641  6.428683  0.000000 13.74008
## N   15.527186 16.065588 13.740077  0.00000

read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]

##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478

```


Index

Phylograms

- dna_phylograms, 10
- immune_phylograms, 10
- rna_phylograms, 11
- snv_heatmaps, 12

Signatures

- anti_pd1_favor_signature, 2
- apoptosis_signature, 3
- drug_resistance_signature, 4
- emt_signature, 5
- hypoxia_signature, 6
- proliferation_signature, 6
- signature_analysis, 8
- specified_geneset_analysis, 7
- z_score_calculations, 2

Multi-Level Divergence and Diversity

- multi_level_divergence_diversity, 13

Utilities

- metaITH_analysis, 14